

EXCEL BASICS FOR DATA SCIENCE



AfterWork

AGENDA

5 MIN

Learning Outcomes

10 MIN

Introduction

120 MIN

Practice

- Icebreaker, Importing and Exporting
- Data Cleaning
- Analysis
- Visualisation

10 MIN

Recap + Feedback Form



AfterWork

LEARNING OUTCOMES

- I can perform **data cleaning** techniques to prepare data for further analysis using Excel.
- I can perform **data analysis** techniques to answer business problems using Excel.
- I can perform **data visualisation** techniques to communicate insights to stakeholders using Excel.

WHAT ARE SPREADSHEETS?

- **Definition**

- Computer programs that capture, display and manipulate data arranged in rows and columns.
- Used to create reports and dashboards which can be comprised of summary tables, visualisations, descriptions, etc.
- Most familiar use is data presentation.
- Formulas are useful for computing sales, budgets, and other numeric summaries.
- Convenient for analysis of structured data and data visualisation.

- **Concepts**

- Cells, Sheets, Formulas, Functions, Charts, Pivot tools, What-if-Analysis etc.

SPREADSHEET APPLICATIONS

- Microsoft Excel by Microsoft (Proprietary)
- Google Sheets by Google (Free)
- Libreoffice Calc (Free and Open Source)
- Lotus 123 by IBM



WHY SPREADSHEETS?

- One of the easiest structured methods to storing and manipulating data.
- Ability to perform statistical calculations.
- Ability to perform data analysis and visualisation
- Best tool to get the job done i.e. producing reports is an easy task.

PROJECTS

- Microsoft Excel is a very versatile tool and can be used for almost anything that you can imagine:

Agendas, Budgets, Calendars, Cards, Charts and diagrams, Financial tools (loan calculators, etc.), Flyers, Forms, Inventories, Invoices, Lists and to-do checklists, Planners, Plans and proposals, Reports, Schedules, Timesheets



ICE BREAKER

Come up with as many questions as possible that can be answered from the clean titanic dataset.

PRACTICE: IMPORTING & EXPORTING

Concepts

- Importing data
- Exporting data

Exercise

- Add a new column to the left of temperature column, Fahrenheit.
- Use formula to compute values from temperature column.
 - Note: Fahrenheit = $1.8C + 32$
- Export the data as a CSV file.

Dataset

- <https://bit.ly/CitiesDataset1>

PRACTICE: DATA CLEANING

Concepts

- Data Cleaning Techniques

Exercise

- Perform the data cleaning techniques outlined in the given guidelines document and save the file for further analysis.
- Data Cleaning guidelines: <https://bit.ly/DataCleaningGuidelines>

Dataset

- <https://bit.ly/DirtyTitanicDataset>

PRACTICE: DATA ANALYSIS

Concepts

- Sorting, Filtering, Aggregation, Joining, Pivot Tables: Restructuring / aggregation / analysis

Exercise

- How many cities in Italy?
- What is the average latitude:
 - Overall?
 - For cities with temperature < 10 ?
 - For cities with temperature > 10 ?
 - For cities where both the city name and the country name end in the letter “a” ?
 - Which are warmer on average – cities in the EU or cities not in the EU?

Dataset

- <https://bit.ly/CitiesDatasetI> and <https://bit.ly/CountriesDatasetI>

PRACTICE: DATA ANALYSIS (TITANIC)

Concepts

- Sorting, Filtering, Aggregation, Joining, Pivot Tables: Restructuring / Aggregation / Analysis.

Exercise

- How many passengers sailed for free?
- How many married women over age 50 embarked in Cherbourg?
- How many passenger ages are missing?
- What is the average fare paid by these passengers?
- What is the most common last name among passengers, and how many passengers have that last name? What is the average number of passengers per last name?
- What is the average fare paid by passengers in the three classes?
- What is the average age of passengers in the three classes?

Dataset

- Clean Titanic Dataset

PRACTICE: DATA VISUALISATION

- Effective Data visualisation
 - Hans Rosling's 200 Countries, 200 Years, 4 Minutes:
 - <https://www.youtube.com/watch?v=jbkSRLYSojo>
- Suggested charts are often good ones.
- For help while working with charts i.e. Web search or fiddling
- Don't underestimate the power of basic visualizations
- Bar charts, Pie charts, Scatterplots and Sparklines

BAR CHARTS

- **Concept**
 - **Bar charts:** When one axis is categories and the other is numeric.
- **Exercise**
 - Ten countries with the highest population, bar chart showing populations.
- **Dataset**
 - <https://bit.ly/CountriesDatasetI>

PIE CHARTS

- **Concept**
 - **Pie charts:** To compare relative sizes of categories.
- **Exercise**
 - Pie chart showing relative number of cities with negative longitude and positive longitude.
- **Dataset**
 - <https://bit.ly/CitiesDatasetI>

SCATTER PLOTS

- **Concept**
 - **Scatter plots:** When both axes are numeric.
- **Exercise**
 - Use a scatter plot to explore whether there is any relationship between the latitude of cities in a country (x-axis) and the population of that country (y-axis).
- **Dataset**
 - <https://bit.ly/CitiesDatasetI> and <https://bit.ly/CountriesDatasetI>

PRACTICE: DATA VISUALISATION (TITANIC)

Concepts

- Data visualisation Techniques: Bar Charts, Pie Charts and Scatter plot

Exercise

- How many passengers sailed for free?
- How many married women over age 50 embarked in Cherbourg?
- How many passenger ages are missing?
- What is the average fare paid by these passengers?
- What is the most common last name among passengers, and how many passengers have that last name? What is the average number of passengers per last name?
- What is the average fare paid by passengers in the three classes?
- What is the average age of passengers in the three classes?

Dataset

- Clean Titanic Dataset

DELIVERABLE



LIMITATIONS

- Automated data cleaning procedures for lots of data can be tedious.
- Manual process for analysis
- Hard to consolidate many data sources through automation
- Lack of version control



DEBRIEF + Q & A

- I can perform **data cleaning** techniques to prepare data for further analysis using Excel.
- I can perform **data analysis** techniques to answer business problems using Excel.
- I can perform **data visualisation** techniques to communicate insights to stakeholders using Excel.





CHECK OUT FORM

<https://bit.ly/ExcelBasicsforDataScience>

GO FORTH AND PROSPER

